3

169.1918

PATENT APPLICATION

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:	)	
	:	Examiner: NYA
JULIE RAE KOWALD	)	
	:	Group Art Unit: 2621
Application No.: 09/730,573	)	
	:	
Filed: December 7, 2000	)	
	:	
For: VISUAL LANGUAGE CLASSIFICATION SYSTEM	)	
	:	April 2, 2001

Commissioner for Patents
Washington, D.C. 20231

CLAIM TO PRIORITY

Sir:

Applicant hereby claims priority under the International Convention and all rights to which he is entitled under 35 U.S.C. § 119 based upon the following Australian Priority Application:

PQ4640 filed December 14, 1999

A certified copy of the priority document is enclosed.

Applicant's undersigned attorney may be reached in

09/730,573

(3)

Patent Office
Canberra

I, CASSANDRA RICHARDS, ACTING TEAM LEADER EXAMINATION SUPPORT & SALES hereby certify that annexed is a true copy of the Provisional specification in connection with Application No. PQ 4640 for a patent by CANON KABUSHIKI KAISHA filed on 14 December 1999.

WITNESS my hand this
Fifteenth day of December 2000

CASSANDRA RICHARDS
ACTING TEAM LEADER
EXAMINATION SUPPORT & SALES

CERTIFIED COPY OF PRIORITY DOCUMENT

S&F Ref: 457300

ORIGINAL

AUSTRALIA

Patents Act 1990

PROVISIONAL SPECIFICATION FOR THE INVENTION ENTITLED:

Emotive Editing System

Name and Address of Applicant:

Canon Kabushiki Kaisha, incorporated in Japan, of 30-2, Shimomaruko 3-chome Ohta-ku, Tokyo, 146, Japan

Name of Inventor:

Julie Rae Kowald

This invention is best described in the following statement:

VISUAL LANGUAGE CLASSIFICATION SYSTEM

Technical Field of the Invention

The present invention relates generally to the classification of image data and, in particular, to a form of automated classification that permits an editor to automatically generate emotive presentations of the image data.

Background

The editing of video of sequences of images (eg. films, video, slide shows), to achieve a desired reaction from an audience traditionally requires input from a human editor who employs techniques other than the mere sequencing of images over a time line. To achieve an understanding by the audience of the intended message or purpose of the production, the editor must draw upon human interpretation methods which are then applied to moving or still images that form the sequence.

Film makers use many techniques to obtain a desired meaning from images, such techniques including the identification and application of different shot types, both moving and still, the use of different camera angles, different lens types and also film effects. The process of obtaining meaning from the images that make up the final production commences with a story or message that is then translated into a storyboard that is used by the film crew and film director as a template. Once the film is captured, the editor is then given the resulting images and a shot list for sequencing. It is at this early stage of production, when the screen writer is translating the written story or script to a storyboard that written language becomes visual language. This occurs due to the method by which the audience is told the story and must interpret the message. The visual nature of a moving image generally only has dialogue relevant to the character's experience and, in most cases, is absent of explicit narrative relative to the story being told and the emotional state of the characters within the story. The screen writers must therefore generate this additional information using the visual language obtained from different shot types.

Examples of different shot types or images are seen in Figs. 1A to 1G. Fig. 1A is representative of an extreme long shot (ELS) which is useful for establishing the characters in their environment, and also orientating the audience as to the particular location. Fig. 1B is representative of a long shot (LS) which is also useful for establishing the characters in their environment and orientating the audience as to the location. In some instances, an ELS is considered more dramatic than the LS. Fig. 1C is representative of a medium long shot (MLS) in which the characters are closer to the viewer and indicates, in a transition from a long shot, subjects of importance to the story.

Typically for human subjects, an MLS views those subjects from the knees upwards. Fig. 1D is indicative of a medium shot (MS) in which human characters are generally shown from the waist upwards, and the shot assists the viewer interpreting the characters reactions to their environment and any particular dialogue taking place. Fig. 1E is indicative of a medium closeup (MCU) in which human characters are generally shown from the chest upwards. The MCU is useful for dialogue and communication interpretation including the emotion of the speaking characters. Fig. 1F is indicative of a closeup (CU) which for human characters frames the forehead and shoulders within the shot, and is useful for clear understanding of the emotions associated with any particular dialogue. The closeup is used to consciously place the audience in the position of the character being imaged to achieve a greater dramatic effect. Fig. 1G is representative of an extreme closeup (ECU) formed by a very tight shot of a portion of the face and demonstrates beyond the dialogue the full dramatic effect of intended emotion. An ECU can be jarring or threatening to the audience in some cases and is often used in many thriller or horror movies. It will further be apparent from the sequence of images in Figs. 1A to 1G that different shots clearly can display different meaning. For example, neither of Figs. 1F and 1G indicate that the subject is seen flying a kite, nor do Figs. 1D or 1E place the kite flying subject on a farm indicated by the cow seen in Figs. 1A to 1C. Further, it is not apparent from Fig. 1A that the subject is smiling or indeed that the subject's eyes are open.

A photograph or moving image of a person incorporating a full body shot will be interpreted by the viewer as having a different meaning to a shot of exactly the same person, where the image consists of only a closeup of the face of the subject. A full-length body shot is typically interpreted by a viewer as informative and is useful to determine the sociological factors of the subject and the relationship of the subject to the particular environment.

An example of this is illustrated in Figs. 2A to 2C which show the same subject matter presented with three different shot types. Fig. 2A is a wide shot of the subject within the landscape and is informative as to the location, subject and activity taken close within the scene. Fig. 2B is a mid-shot of the subject with some of the surrounding landscape, and changes the emphasis from the location and activity to the character of the subject. Fig. 2C provides a closeup of the subject and draws the audience to focus upon the subject.

Panning is a technique used by screen writers to help the audience participate in the absorption of information within a scene. The technique is commonly used with open

landscapes or when establishing shots are used in movie productions. A straight shot, obtained when the camera does not move, contrasts the effectiveness of a pan. With a straight shot, the viewer is forced to move their eyes around the scene, searching for information, as opposed to how the pan feeds information to the viewer thus not requiring the viewer to seek out a particular message. The movement of the camera within a pan directs the audience as to those elements within a scene that should be observed and, when used correctly, is intended to mimic the human method of information interpretation and absorption. Fig. 3A is an example of a still shot including a number of image elements (eg. the sun, the house, the cow, the person and the kite) which the audience may scan for information. In film, a still shot is typically used as an establishing shot so as to orientate the audience with the location and the relationship to the story. The screen writer relies upon this type of shot to make sense of any following scenes. Fig. 3B demonstrates an example of a panning technique combined with a zoom, spread amongst four consecutive frames.

Further, differing camera angles, as opposed to direct, straight shots, are often used to generate meaning from the subject, such meaning not otherwise being available due to dialogue alone. For example, newspaper and television journalists often use altered camera angles to solicit propaganda about preferred election candidates. For example, interviews recorded from a low angle present the subject as superior to the audience, whereas the presentation of the same subject may be altered if taken from a high angle to give an inferior interpretation. The same technique is commonly used in movie making to dramatically increase the effect of an antagonist and their victim. When the victim is shot from a high angle, they not only appear as weak and vulnerable, but the audience empathises with the character also experiences their fear.

Fig. 4A is indicative of an eye level shot which is a standard shot contrasting with angles used in other shots and seen in Figs. 4B to 4E. Fig. 4B shows a high angle shot and is used to place the subject in an inferior position. Fig. 4C is indicative of a low angle shot where the camera angle is held low with the subject projecting them as superior. Fig. 4D is indicative of an oblique angle shot where the camera is held off-centre influencing the audience to interpret the subject as out of the ordinary, or as unbalanced in character. Fig. 4E is representative of a Dutch angle shot which is often used to generate a hurried, "no time to waste" or bizarre effect of the subject. The audience is conveyed a message that something has gone astray in either a positive or negative fashion.

There are many other types of images or shots in addition to those discussed above that can give insight to the particular story being presented. Tracking shots follow the subject allowing the audience the experience of being part of the action. Panning gives meaning and designates importance to subjects within a scene as well as providing a panoramic view of the scene. A "swish" pan is similar however is used more as a transition within a scene, quickly sweeping from one subject to another, thus generating a blurred effect. Tilt shots consist of moving the camera from one point up or down, thus mimicking the way in which humans evaluate a person or vertical object absorbing the information presented thereby. A hand-held shot portrays to the audience that the filming is taking place immediately, and if often used to best effect when associated with shots taken when the camera is supported (eg. using a tripod or boom).

To understand the impact visual language has on presenting images in a more meaningful way, it is appropriate to compare the results of contemporary motion pictures with earlier attempts of film making. Early examples of motion pictures consisted of full shots of the characters from the feet upwards reflecting the transition from stage acting. For example, the Charlie Chaplin era of film making and story telling contrasts sharply with later dramatic, emotion filled motion pictures. D.W. Griffiths first notably introduced the use of a pallet of shot types for the purpose of creating drama in film. This arose from a desire of the audience to explore the emotional experience of the characters of the film.

Film makers also use other techniques to tell their story, such techniques including the choice of lens and film effects. These are all used to encourage the audience to understand the intended message or purpose of the production. The audience does not need to understand how, or even be aware that, these techniques have been applied to the images. In fact, if applied properly with skill, the methods will not even be apparent to the audience.

The skill required by the successful film maker is typically only acquired through many years of tuition and practice as well as through the collaboration of many experts to achieve a successfully crafted message. Amateur film makers and home video makers in contrast often lack the skill and the opportunity to understand or employ such methods. However, amateur and home film makers, being well exposed to professional film productions have a desire for their own productions to be refined to some extent approaching that of professional productions, if not those of big-budget Hollywood extravaganzas. Whilst there currently exists many film schools that specialise in courses to educate potential film makers with such techniques, attendance at such courses is often

prohibitive to the amateur film maker. Other techniques currently available that may assist the amateur film maker typically includes software products to aid in the sequencing of images and/or interactive education techniques for tutoring prospective film makers. However, current software approaches have not been widely adopted due to prohibitive costs and skill required for use being excessive for small (domestic) productions.

Time is also a major factor in respect to the current techniques of film editing to unskilled editor. Typically, the time taken to plan shots and their sequencing is substantial and is typically out of the realistic scope of an average home/amateur film maker.

It is therefore desirable to provide a means by which unskilled (amateur) movie makers can create visual productions that convey a desired emotive effect to an audience without a need for extensive planning or examination of shot types.

Disclosure of the Invention

The present invention acts to address this need through the automated classification of images and/or shots into various emotive categories thereby permitting editing to achieve a desired emotive effect.

According to a first aspect of the invention, there is provided a method for automated classification of a digital image, said method comprising the steps of:

analysing said image for the presence of a human face;

determining a size of the located face with respect to a size of said image; and

classifying said image based on the relative size of said face with respect to said image.

According to a second aspect of the invention, there is provided a method for automated classification of a digital image, said method comprising the steps of:

analysing said image for the presence of a human face;

determining a position of the located face with respect to a frame of said image; and

classifying said image based on the relative position of said face with respect to said image frame.

According to another aspect of the invention, there is provided an apparatus for implementing any one of the aforementioned methods.

According to another aspect of the invention there is provided a computer program product including a computer readable medium having recorded thereon a computer program for implementing any one of the methods described above.

Brief Description of the Drawings

A number of preferred embodiments of the present invention will now be described with reference to the drawings, in which:

Figs. 1A to 1G depict a number of shot ranges used by film makers;

Figs. 2A to 2C depict three different shot types used by film makers;

Figs. 3A and 3B depict the effect of a pan in influencing the emotional state of the viewer;

Figs. 4A to 4E depict various angled camera shots also used by film makers;

Fig. 5 is a schematic block diagram representation of a system incorporating the preferred embodiment; and

Fig. 6 is a schematic block diagram of a general purpose computer system upon which the preferred embodiment of the present invention can be practiced.

Detailed Description including Best Mode

Fig. 5 shows a schematic representation of an image recording and production system 500 according to the preferred embodiment where a scene 502 is captured using an image recording device 504, such as a digital video camera or digital still camera. When the scene 502 is captured by a still camera, typically a sequence of still images is recorded, in effect complementing the sequence of images that might be recorded by a video camera. Associated with the capture of the images is the generation of capture data 506 which is output from the camera 504 and typically comprises image data, video data, audio data and metadata.

Where appropriate, the capture data 504 recorded by the camera 504 is transferred 508 to a mass storage arrangement 510, typically associated with a computing system, whereupon the images are made available via an interconnection 520 to a visual language classification system 522. The classification system 508 generates metadata which is configured for convenient editing by the film maker. The visual language classification system 522 outputs classification data 524, configured as further metadata, which is associated with each image and which may be stored within a mass storage unit 526. The classification data 524 in the store 526 may be output to an editing module 514 which, through accessing the image data via a connection 512 to the store 510, provides for the formation of an edited sequence 528 which may be output to a presentation unit 516 for display via a display unit 518, such as a television display, or storage in a mass storage device 519. In some implementations, the stores 510, 526 and 519 may be integrally formed.

The classification system 522 performs content analysis to analyse the images residing in the store 510. The analysis performed within the classification system 522 is configured to provide information about the intention of the photographer at the time of capturing the image or image sequence. Such analysis comprises the detection of human faces and preferably other visually distinct features including landscape features such as the sky, green grass, sandy or brown earth, or other particular shapes such as motor vehicles, buildings and the like. Audio analysis where appropriate can be used to identify specific events within the sequence of images such a person talking, the passing of a motor car, or the crack of a ball hitting a bat in a sports game, such as baseball or cricket, for example. The classification system 522 provides metadata related to or indicative of the content identified within an image sequence, or at the particular image within the sequence.

One specific example of content analysis that may be applied by classification system 522 is that of face detection, that permits identification and tracking of particular human subjects in images or sequences thereof. An example of a face detection arrangement that may be used in the preferred embodiment is that described in US Patent No. 5,642,431-A (Poggio et. al.). Another example is that disclosed in Australian Patent Publication No. AU-A-33982/99 corresponding to United States Patent Application No. 09/326,561 (Lennon et. al.) (Ref's: CFP1327AU IPR20 461584). Such face detection arrangements typically identify within an image frame a group or area of pixels which are skin coloured and thus may represent a face, thereby enabling that group or area, and thus the face, to be tagged by metadata and monitored. Such monitoring may include establishing a bounding box about the height and width of the detected face and thereafter tracking changes or movement in the box across a number of image frames.

In the sequence of images of Figs. 1A to 1G, the fine content of Figs. 1A and 1B are generally too small to permit accurate face detection. As such, those frames may be classified as non-face images. However in each of Figs. 1C to 1G, the face of the person flying the kite is quite discernible and a significant feature of each respective image. Thus, those images may be automatically classified as face images, such classification being identified as metadata generated by the content analysis unit 610 and linked or otherwise associated with the metadata 612 provided with the images.

Further, and according to the preferred embodiment, the size of the face as a proportion of the overall image size is used to establish and record the type of shot. For example, simple rules may be established to identify the type of shot. A first rule can be that where a face is detected, but the face is substantially smaller than the image in which

the face is detected, that image may be classified as a far shot. A similar rule is where a face is detected which is sized substantially the same as the image. This may be classified as a close-up. An extreme close-up may be where the face occupies the entire image or where it is substantially the same size as the image but extends beyond the edges of the image.

In another example, in Fig. 1C, which is a MLS, the face represents about 2% of the image. In Fig. 1D, the face occupies about 4% of the image, this being a MS. For Fig. 1E, a MCU delivers the face at a size of about 10% of the image. The CU shot of Fig. 1F provides the face at about 60% of the image, and for a ECU, the face is about 60% of the image. A suitable set of rules may thus be established to define the type of shot relative to the subject, whether or not the subject is a face or some other identifiable image structure (eg. cow, house, motor vehicle, etc). Example rules are set out below:

Medium Long Shot (MLS)	subject < 2.5% of the image;
Medium Shot (MS)	2.5% < subject < 10% of the image;
Medium Close Up (MCU)	10% < subject < 30% of the image;
Close Up	30% < subject < 80% of the image; and
Extreme Close Up	subject > 80% of the image.

Where desired, the film maker may vary the rule depending on the particular type of source footage available, or depending on a particular editing effect desired to be achieved.

Another example of content analysis for classification is camera tilt angle. This can be assessed by examining the relative position of a detected face in the image frame. For example, as seen in Fig. 4A, where the face is detected centrally within the image frame, this may be classified as a eye-level shot. In Fig. 4B, where the subject is positioned towards the bottom of the frame, such may be classified as a high angle shot. the positioning of the detected face may be correlated with a tiling of the image frame so as to provide the desired classification. Tiles within the frame may be pre-classified as eye-level, high shot, low shot, left side, and right side. The location of the detected face in certain tiles may then be used to determine an average tile location and thus classify the image according to the position of the average face tile. Such an approach may be readily applied to the images of Figs. 4A to 4D.

The Dutch shot of Fig. 4E may be determined by detecting edges within the image. Such edges may be detected using any one of a large number of known edge detection arrangements. Edges in images often indicate the horizon, or some other horizontal edge, or vertical edges such as those formed by building walls. An edge that is

detected as being substantially non-vertical and non-horizontal may thus indicate a Dutch shot. Classification may be performed by comparing an angle of inclination of the detected edge with the image frame. Where the angle is about 0 degrees or about 90 degrees, such may be indicative of an horizon or vertical wall respectively. Such may be a traditional shot. However, where the angle of inclination is substantially between these values, a Dutch shot may be indicated. Preferred angles of inclination for such detection may be between 30 and 60 degrees, but may be determined by the user where desired.

In an alternative embodiment, the visual language classification system can permit the user to supplement the classification with other terms relating to the emotive message conveyed by the scene. Such manually entered metadata may include terms such as "happy", "smiling", "leisure", and "fun" in the example of Figs. 1C to 1G. More complicated descriptions may also be entered, such as "kite flying". This manually enter metadata that can supplement the automatically generate metadata and be stored with the automatically generated metadata.

As a result of such processing, the store 526 is formed to include metadata representative of the content of source images to be used to form the final production. The metadata not only includes timing and sequencing (eg. scene number etc.) information, but also information indicative of the content of the images and shot types which can be used as prompts in the editing process to follow.

With the database 526 formed, the user may then commence editing the selected images. This is done by invoking an editing system 514 which extracts the appropriate images or sequence of images from the store 510. Using the information contained within the metadata store 526, the user may conveniently edit particular images. The database information may be used to define fade-in and fade-out points, images where a change in zoom is desired, points of interest within individual images which can represent focal centres for zooming operations either or both as source or target, amongst many others.

Editing performed by the editing system 514 may operate using the classifications 524 in a variety of ways. For example, the user may wish to commence an image sequence with a long shot, and hence may enter into the system 514 a request for all long shots to be listed. The system 514 then interrogates the store 526 to for a pick-list of images that have been previously classified as a long shot. The user may then select a long shot from the list to commence the edited sequence. The classification thus substantially reduces the user's editing time by providing a ready source of searchable information regarding each image or shot sequence. Another example is where the user wishes to show fear in the faces of the subjects. Since faces are typically not detected in

any significant detail for anything under a medium shot, a search of the store 526 may be made for all medium shots, close-ups and extreme close-ups. A corresponding pick list results from which the user can conveniently review a generally smaller number of images than the total number available to determine those that show fear. User entered metadata such as "fear" may then supplement the automatically generated classification for those images that display such an emotion.

The automated content analysis of images as discussed above permits the rapid processing of sequences of image to facilitate the formation of an enhanced edited result. For example, where a video source is provide having 25 frames per second, a 5 second shot requires the editing of 125 frames. To perform manual face detection and focal point establishment on each frame is time consuming and prone to inconsistent results due to human inconsistency. Through automation by content analysis, the positions of the face since each frame may be located according to consistently applied rules. All that is then necessary is form the user to select the start and end points and the corresponding edit functions (eg. zoom values from. 0% at the start, and 60% at the end).

Metadata analysis of the source material may include the following:

(i) time code and date data;

(ii) GPS data;

(iii) image quality analysis (sharpness, colour, content quality, etc.);

(iv) original shot type detection;

(v) object detection and custom object detection (determined by the author);

(vi) movement detection;

(vii) face detection;

(viii) audio detection;

(ix) collision detection;

(x) tile (interframe structure) analysis; and

(xi) User entered metadata.

The method described above with reference to Fig. 5 is preferably practiced using a conventional general-purpose computer system 600, such as that shown in Fig. 6 wherein the processes of Fig. 5 may be implemented as software, such as an application program executing within the computer system 600. The software may be divided into two separate parts; one part for carrying out the classification and editing methods, and another part to manage the user interface between the latter and the user. The software may be stored in a computer readable medium, including the storage devices described

below, for example. The software is loaded into the computer from the computer readable medium, and then executed by the computer. A computer readable medium having such software or computer program recorded on it is a computer program product. The use of the computer program product in the computer preferably effects an advantageous apparatus for classification and consequential editing in accordance with the embodiments of the invention.

The computer system 600 comprises a computer module 601, input devices such as a keyboard 602 and mouse 603, output devices including a printer 615 and a visual display device 614 and loud speaker 617. A Modulator-Demodulator (Modem) transceiver device 616 is used by the computer module 601 for communicating to and from a communications network 620, for example connectable via a telephone line 621 or other functional medium. The modem 616 can be used to obtain access to the Internet, and other network systems, such as a Local Area Network (LAN) or a Wide Area Network (WAN).

The computer module 601 typically includes at least one processor unit 605, a memory unit 606, for example formed from semiconductor random access memory (RAM) and read only memory (ROM), input/output (I/O) interfaces including a audio/video interface 607, and an I/O interface 613 for the keyboard 602 and mouse 603 and optionally a joystick (not illustrated), and an interface 608 for the modem 616. A storage device 609 is provided and typically includes a hard disk drive 610 and a floppy disk drive 611. A magnetic tape drive (not illustrated) may also be used. A CD-ROM drive 612 is typically provided as a non-volatile source of data. The components 605 to 613 of the computer module 601, typically communicate via an interconnected bus 604 and in a manner which results in a conventional mode of operation of the computer system 600 known to those in the relevant art. Examples of computers on which the embodiments can be practised include IBM-PC's and compatibles, Sun Sparcstations or alike computer systems evolved therefrom.

Typically, the application program of the preferred embodiment is resident on the hard disk drive 610 and read and controlled in its execution by the processor 605. Intermediate storage of the program and any data fetched from the network 620 may be accomplished using the semiconductor memory 606, possibly in concert with the hard disk drive 610. In some instances, the application program may be supplied to the user encoded on a CD-ROM or floppy disk and read via the corresponding drive 612 or 611, or alternatively may be read by the user from the network 620 via the modem device 616. Still further, the software can also be loaded into the computer system 600 from other

computer readable medium including magnetic tape, a ROM or integrated circuit, a magneto-optical disk, a radio or infra-red transmission channel between the computer module 601 and another device, a computer readable card such as a PCMCIA card, and the Internet and Intranets including e-mail transmissions and information recorded on Websites and the like. The foregoing is merely exemplary of relevant computer readable mediums. Other computer readable mediums may be practiced without departing from the scope and spirit of the invention.

The method of described with reference to Fig. 6 may alternatively or additionally be implemented in dedicated hardware such as one or more integrated circuits performing the functions or sub functions of the system. Such dedicated hardware may include graphic processors, digital signal processors, or one or more microprocessors and associated memories. For example, specific visual effects such as zoom and image interpolation may be performed in specific hardware devices configured for such functions. Other processing modules, for example, used for face detection or audio processing, may be performed in dedicated DSP apparatus.

Industrial Applicability

Embodiments of the invention are applicable to the image editing and reproduction industries and find particular application with amateur movie makers who are trained in the intricacies of shot and subject identification, and consequential editing based thereupon..

The foregoing describes only some embodiments of the present invention, and modifications and/or changes can be made thereto without departing from the scope and spirit of the present invention, the described embodiments being illustrative and not restrictive.

In the context of this specification, the word "comprising" means "including principally but not necessarily solely" or "having" or "including" and not "consisting only of". Variations of the word comprising, such as "comprise" and "comprises" have corresponding meanings.

Claims:

1. A method for automated classification of a digital image, said method comprising the steps of:

analysing said image for the presence of a human face;

determining a size of the located face with respect to a size of said image; and

classifying said image based on the relative size of said face with respect to said image.

2. A method according to claim 1 wherein said image is classified using a term which provides information about an intention of a photographer whom captured said image.

3. A method according to claim 1 or 2 wherein said image is classified as a far-shot if the size of said located face is substantially less than the size of said image.

4. A method according to claim 1or 2 wherein said image is classified as a close-up where the size of said located face substantially corresponds with the size of said image

5. A method according to claim 1 or 2 wherein said image is classified as an extreme close-up where only a part of said located face appears within said image.

6. A method according to claim 1 or 2 wherein said classifying comprises associating a size of said located face with a set of predetermined thresholds for a size of a human face image.

7. A method according to claim 1 or 2 wherein said image is classified as a far shot if said image contains a face and the size of said located face is below a first predetermined threshold compared to the size of said image.

8. A method according to claim 7 wherein said image is classified as an extreme close up if the size of said located face is above a second predetermined threshold compared to the size of said image.

9. A method according to claim 8 wherein said image is classified as a close-up if the size of said located face is below said second predetermined threshold and above a third predetermined threshold compared to the size of said image.

10. A method according to claim 9 wherein said image is classified is a medium shot if the size of said located face is greater than said first predetermined threshold and less than said third predetermined threshold.

11. A method according to any one of the preceding claims wherein said analysing comprises information provided with said image.

12. A method according to any one of the preceding claims wherein said image comprises a frame of digital video sequence of images.

13. A method according to claim 12 wherein said information is associated with other frames of said sequence.

14. A method according to any one of the preceding claims wherein said analysing comprises detecting one or more regions of said image at which skin coloured pixels are located in order to locate said face.

15. A method according to any one of the preceding claims wherein said determining approximates the size of said located face by a height and width of a bounding rectangle that encloses said face.

16. A method for automated classification of a digital image, said method comprising the steps of:

analysing said image for the presence of a human face;

determining a position of the located face with respect to a frame of said image; and

classifying said image based on the relative position of said face with respect to said image frame.

17. A method according to claim 16 wherein said image is classified using a term which provides information about an intention of a photographer whom captured said image.

18. A method according to claim 16 or 17 wherein said image is classified as a high-shot if the position of said located face is substantially toward a bottom of said image frame.

19. A method according to claim 16 or 17 wherein said image is classified as a eye-level shot where the position of said located face substantially corresponds with a centre of said image frame.

20. A method according to claim 16 or 17 wherein said image is classified as a low shot where the position of said located face is substantially toward a top of said image frame.

21. A method according to claim 16 or 17 wherein said image is classified as a left shot where the position of said located face is substantially toward a right hand side of said image frame.

22. A method according to claim 16 or 17 wherein said image is classified as a right shot where the position of said located face is substantially toward a left hand side of said image frame.

23. A method according to claim 16 or 17 wherein said image is classified as a low shot where the position of said located face is substantially toward a top of said image frame.

24. A method according to any one of claims 16 to 23 wherein said analysing comprises information provided with said image.

25. A method according to any one of claims 16 to 24 wherein said image comprises a frame of digital video sequence of images.

26. A method according to claim 25 wherein said information is associated with other frames of said sequence.

27. A method according to any one of claims 1 to 26 further comprising the steps of:

detecting an edge within said image;

determining an angle of inclination between said edge and an axis of said image frame;

classifying said image as a Dutch shot where said angle of inclination is between predetermined angles of inclination.

28. A method according to claim 27 wherein said predetermined angles of inclination comprise 30 and 60 degrees.

29. A method for automated classification of a digital image substantially as described herein with reference to any one of the embodiments of the method as that embodiment is illustrated in the drawings.

30. A method of processing at least one image, said method comprising the steps of:

classifying said at least one image using a method according to any one of claims 1 to 29; and

editing said at least one image using said classification to form a sequence of edited images.

31. A system for performing the method of any one of the preceding claims.

32. A computer program product incorporating a computer readable medium having a series of instructions form performing a method according to any one of claims 1 to 16.

33. An edited sequence of images formed through implementation of a method according to claim 30.

Dated this fourteenth Day of December, 1999
Canon Kabushiki Kaisha
Patent Attorneys for the Applicant/Nominated Person
SPRUSON & FERGUSON

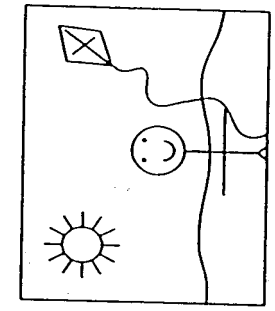

FIG. 1D

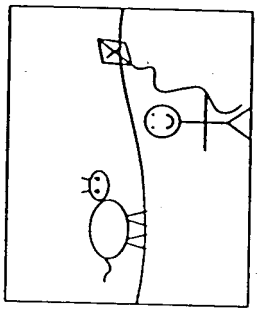

FIG. 1C

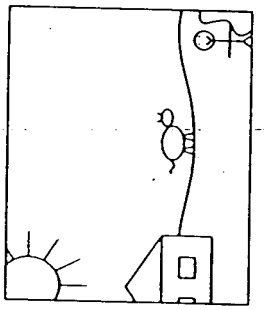

FIG. 1B

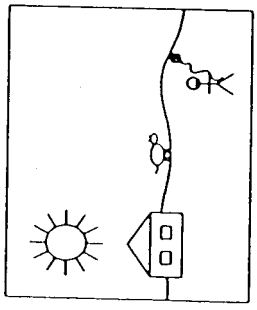

FIG. 1A

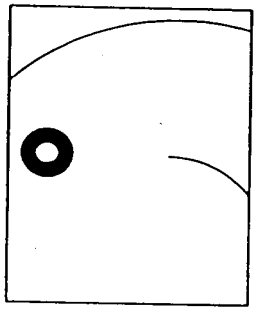

FIG. 1G

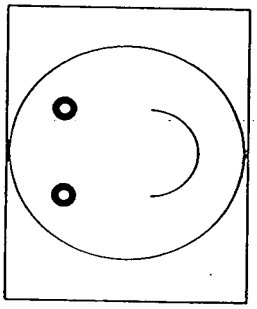

FIG. 1F

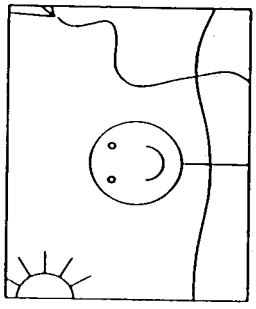

FIG. 1E

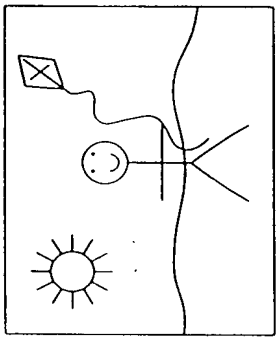

FIG. 2A

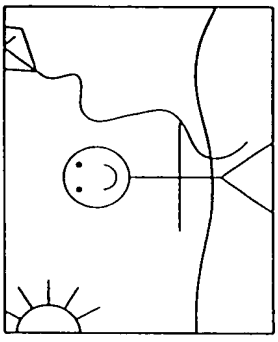

FIG. 2B

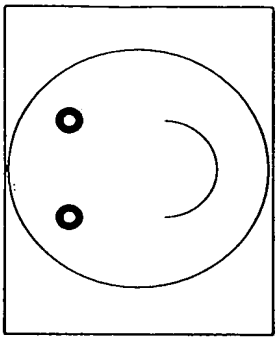

FIG. 2C

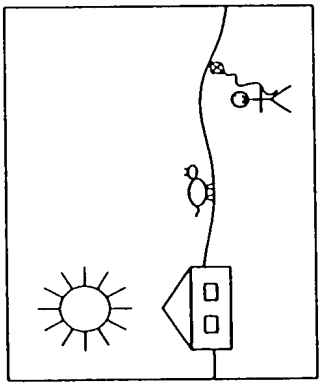

FIG. 3A

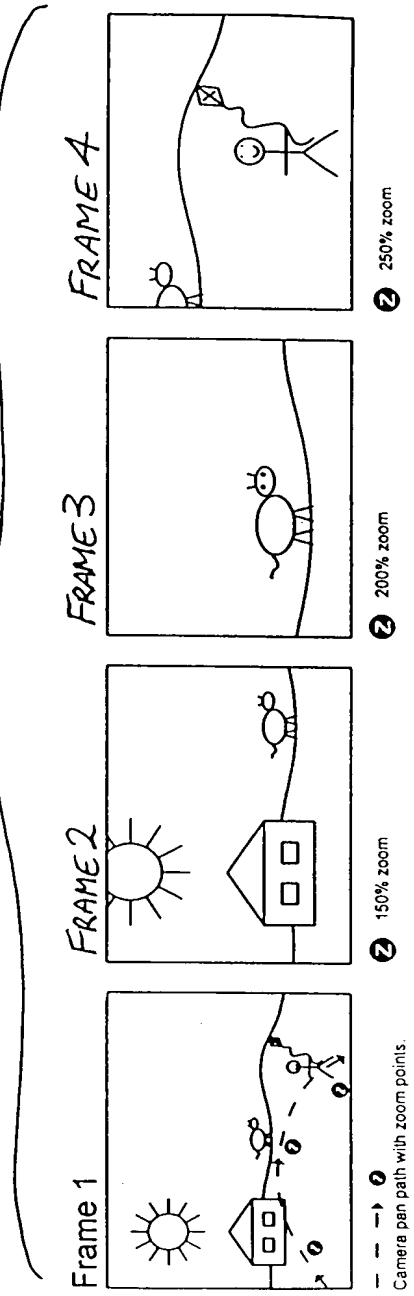

FIG. 3B

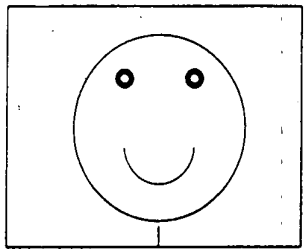

FIG. 4A

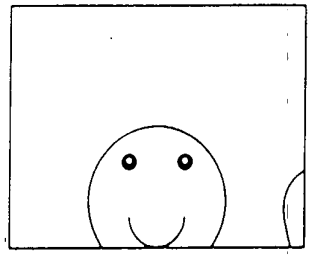

FIG. 4B

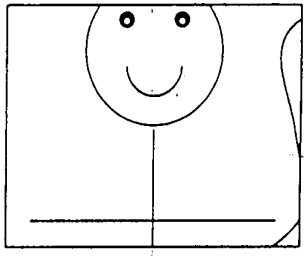

FIG. 4C

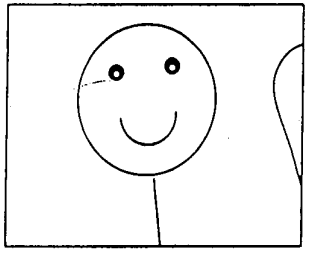

FIG. 4D

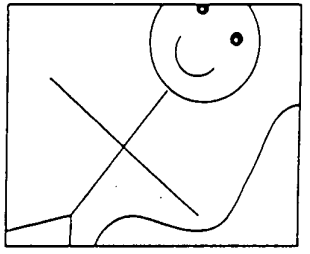

FIG. 4E

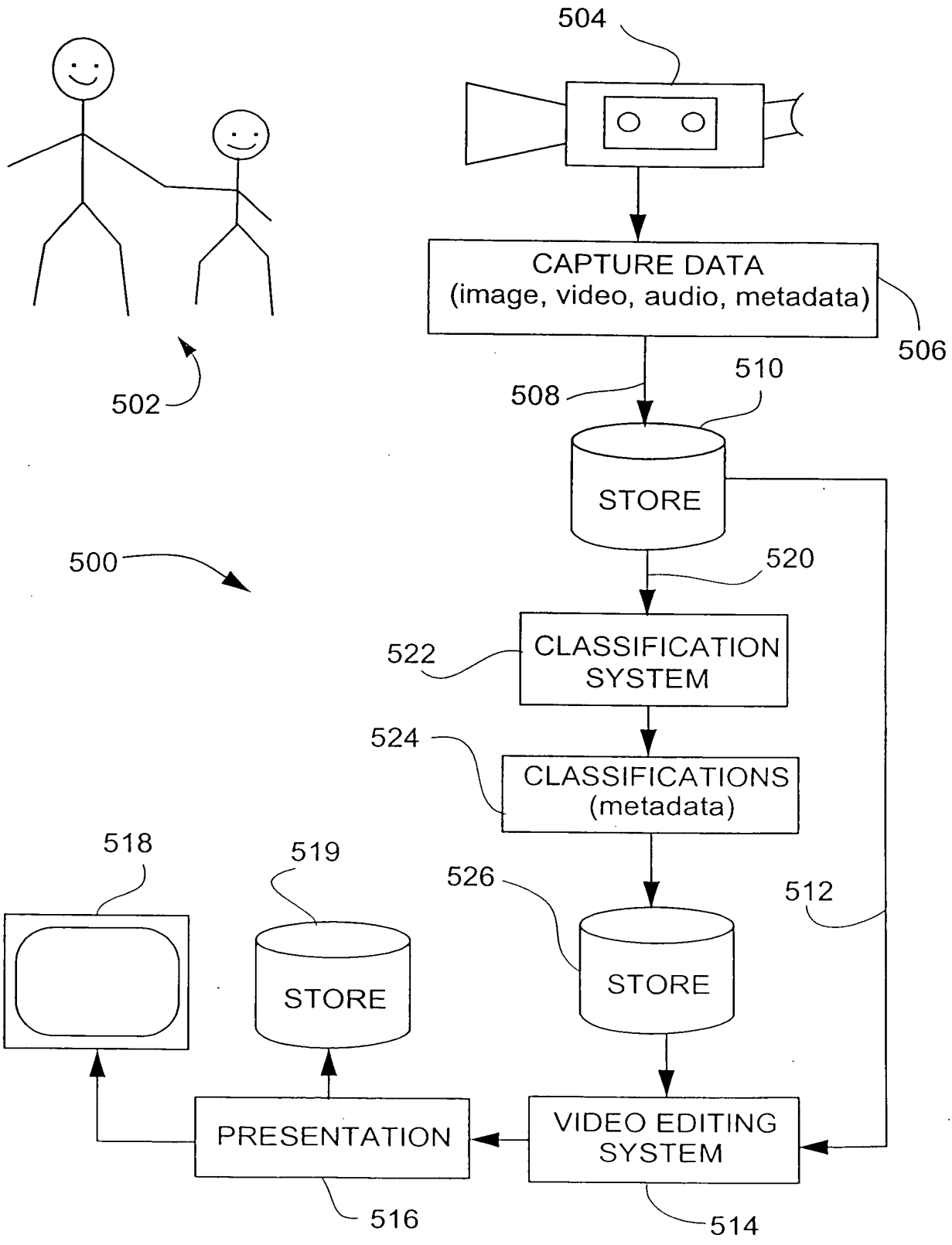
504
CAPTURE DATA
(image, video, audio, metadata)
506
502
508
510
STORE
500
520
522
CLASSIFICATION SYSTEM
524
CLASSIFICATIONS
(metadata)
518
519
526
512
STORE
STORE
PRESENTATION
VIDEO EDITING SYSTEM
516
514

Fig. 5

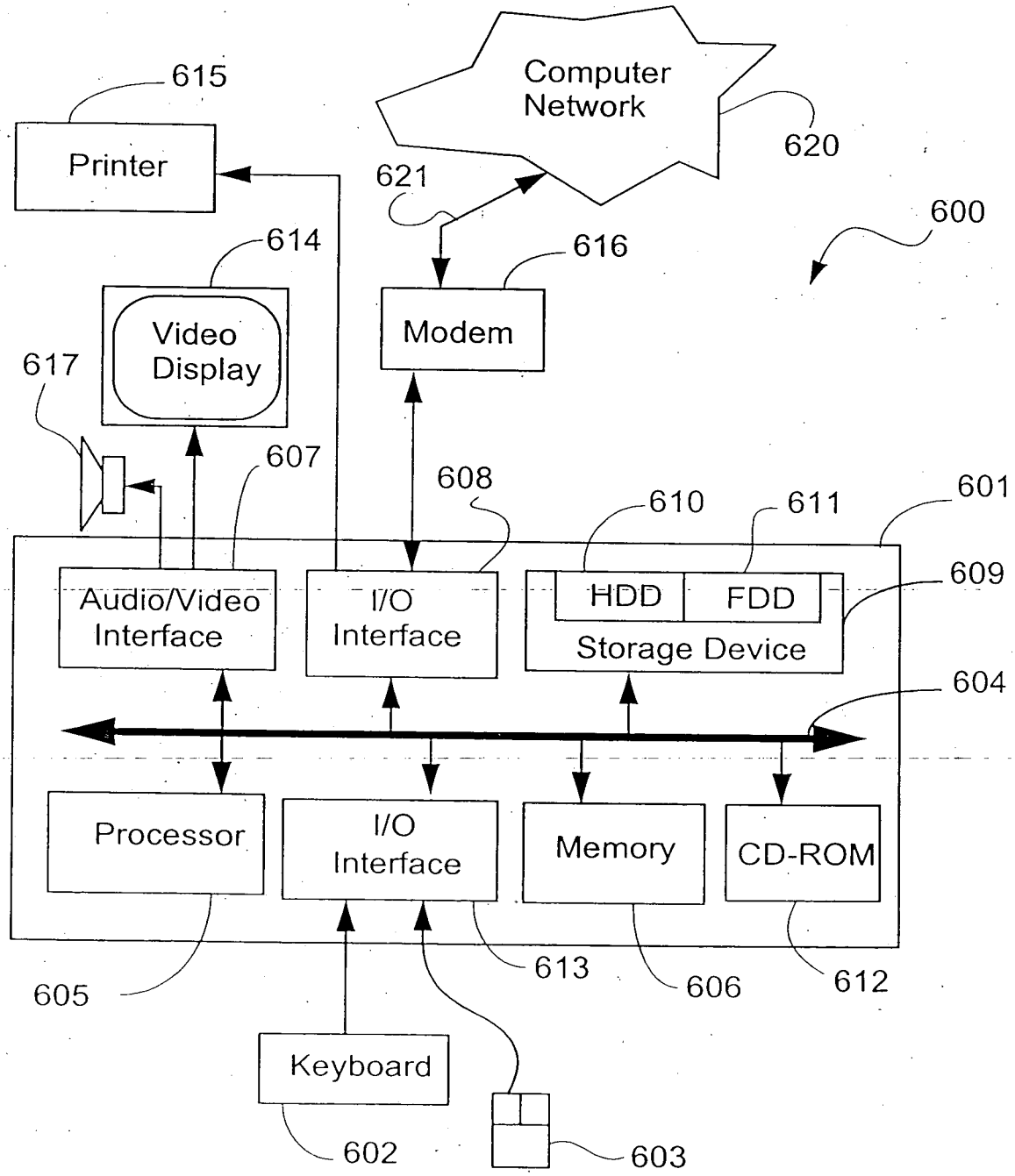
615
Printer
Computer Network
620
621
600
614
616
Video Display
Modem
617
607
608
610
611
601
609
Audio/Video Interface
I/O Interface
HDD
FDD
Storage Device
604
Processor
I/O Interface
Memory
CD-ROM
605
613
606
612
Keyboard
602
603

Fig. 6